

# TABLA DE CONTENIDOS

|   | página     |
|---|------------|
| <b>AGRADECIMIENTOS</b>  | <b>II</b>  |
| <b>TABLA DE CONTENIDOS</b>                                    | <b>III</b> |
| <b>ÍNDICE DE FIGURAS</b>                                      | <b>VI</b>  |
| <b>ÍNDICE DE TABLAS</b>                                       | <b>VII</b> |
| <b>RESUMEN</b>  | <b>IX</b>  |
| <b>ABSTRACT</b>   | <b>XI</b>  |
| <b>1. INTRODUCCIÓN</b>  | <b>1</b>   |
| 1.1. Marco Teórico . . . . .                                  | 1          |
| 1.1.1. <i>Vibrio parahaemolyticus</i> . . . . .               | 1          |
| 1.1.1.1. <i>V. parahaemolyticus</i> pandémico O3:K6 . . . . . | 2          |
| 1.1.2. Small-RNA . . . . .                                    | 3          |
| 1.1.3. Next-Generation Sequencing (NGS) . . . . .             | 7          |
| 1.1.3.1. Preparación Biblioteca mRNA para Illumina . . . . .  | 8          |
| 1.1.3.2. Secuenciación Illumina . . . . .                     | 9          |
| 1.1.3.3. RNA-seq . . . . .                                    | 12         |
| 1.2. Presentación del Problema . . . . .                      | 13         |
| 1.3. Justificación del Problema . . . . .                     | 14         |
| <b>2. HIPÓTESIS</b>   | <b>15</b>  |
| 2.1. Hipótesis . . . . .                                      | 15         |
| <b>3. OBJETIVOS</b>   | <b>16</b>  |
| 3.1. Objetivo General . . . . .                               | 16         |
| 3.2. Objetivos Específicos . . . . .                          | 16         |
| <b>4. MATERIALES y MÉTODOS</b>                                | <b>17</b>  |
| 4.1. Cultivo de bacterias y obtención de RNA . . . . .        | 18         |
| 4.2. Preparación de biblioteca . . . . .                      | 19         |
| 4.3. Análisis de calidad . . . . .                            | 20         |

|           |  |           |
|-----------|--|-----------|
| 4.4.      | Selección aleatoria de lecturas (Random Sampling)  | 21        |
| 4.5.      | Filtrado de lecturas   | 22        |
| 4.6.      | Alineamiento de lecturas   | 23        |
| 4.7.      | Creación de formatos BAM, ordenamiento y creación de index de archivos<br>formato BAM                                  | 25        |
| 4.8.      | Visualización de alineamientos (IGV/Artemis)   | 26        |
| 4.9.      | Creación de archivo en Formato General de Transferencia (GTF)  | 27        |
| 4.10.     | Conteo de lecturas   | 28        |
| 4.11.     | Análisis estadístico de conteos  | 30        |
| 4.12.     | Matriz de Conteo   | 31        |
| 4.13.     | Expresión Diferencial  | 31        |
| 4.14.     | Análisis Biológico   | 33        |
| <b>5.</b> | <b>RESULTADOS</b>  | <b>34</b> |
| 5.1.      | Análisis de calidad  | 34        |
| 5.2.      | Selección aleatoria de lecturas (Random Sampling)  | 36        |
| 5.3.      | Filtrado de lecturas   | 36        |
| 5.4.      | Alineamiento de lecturas al genoma de <i>V. parahaemolyticus</i> y Visualización<br>de los alineamientos (IGV/Artemis) | 37        |
| 5.5.      | Conteo de lecturas correspondientes a los distintos sRNA   | 39        |
| 5.6.      | Abundancia Total de Transcritos  | 39        |
| 5.7.      | Reproducibilidad de las muestras y Expresión relativa de los sRNA  | 41        |
| 5.8.      | Expresión Diferencial  | 43        |
| 5.9.      | Regulación de los sRNA altamente expresados  | 44        |
| 5.10.     | Análisis Biológico   | 45        |
| <b>6.</b> | <b>DISCUSIÓN</b>   | <b>47</b> |
| 6.1.      | Análisis de calidad  | 47        |
| 6.2.      | Selección aleatoria de lecturas (Random Sampling)  | 48        |
| 6.3.      | Filtrado de lecturas   | 48        |
| 6.4.      | Comparación de Herramientas  | 49        |
| 6.4.1.    | Comparación Herramientas de Alineamiento (BWA y SMALT)   | 49        |
| 6.4.2.    | Comparación Herramientas de Visualización (IGV y Artemis)  | 50        |
| 6.5.      | Abundancia Total de Transcrito   | 51        |
| 6.6.      | Reproducibilidad de las muestras y Expresión relativa de los sRNA  | 53        |
| 6.7.      | Expresión Diferencial y Regulación de los sRNA altamente expresados  | 53        |

|   |           |
|---|-----------|
| <b>7. CONCLUSIÓN</b>  | <b>59</b> |
| <b>8. BIBLIOGRAFÍA</b>                                      | <b>61</b> |
| <b>9. ANEXOS</b>  | <b>68</b> |
| 9.1. FORMATO FASTQ . . . . .                                | 68        |
| 9.2. FORMATO SAM . . . . .                                  | 69        |
| 9.3. COMANDOS METODOLOGÍA . . . . .                         | 70        |
| 9.4. ARCHIVO GTF CREADO . . . . .                           | 71        |
| 9.5. MATRIZ DE CONTEO . . . . .                             | 72        |
| 9.6. RESULTADOS EXTRA . . . . .                             | 75        |
| 9.7. SCRIPT EN PYTHON PARA CALCULAR LOS RPKM DE LOS sRNA .  | 84        |
| 9.8. SCRIPT EN R PARA CALCULAR LA EXPRESIÓN DIFERENCIAL . . | 87        |
| 9.9. LISTA DE GENES EXPRESADOS DIFERENCIALMENTE . . . . .   | 89        |

## ÍNDICE DE FIGURAS

|  | página |
|--|--------|
| 1.1. Preparación de Bibliotecas para mRNA Illumina . . . . .           | 9      |
| 1.2. Secuenciación Illumina . . . . .                                  | 11     |
| 4.1. Diagrama conceptual flujo de trabajo . . . . .                    | 17     |
| 4.2. Modos de conteo que realiza HTSeq-count . . . . .                 | 30     |
| 5.1. Análisis de Calidad muestra Aislada . . . . .                     | 35     |
| 5.2. Análisis de Calidad muestra Ambiental . . . . .                   | 35     |
| 5.3. Visualización con IGV Resultados Alineamiento BWA . . . . .       | 38     |
| 5.4. Visualización con IGV Resultados Alineamiento SMALT . . . . .     | 38     |
| 5.5. Visualización con Artemis Resultados Alineamiento BWA . . . . .   | 39     |
| 5.6. Visualización con Artemis Resultados Alineamiento SMALT . . . . . | 39     |
| 5.7. HTSeq-count: -a=10, -m=union . . . . .                            | 40     |
| 9.1. Archivo en formato fastq . . . . .                                | 68     |
| 9.2. HTSeq-count: -a=10, -m=union . . . . .                            | 75     |

## ÍNDICE DE TABLAS

|  | página |
|--|--------|
| 1.1. Características Generales de <i>V. parahaemolyticus</i> . . . . .   | 3      |
| 4.1. Condiciones estudiadas para <i>V. parahaemolyticus</i> . . . . .  | 18     |
| 4.2. Cantidad de lecturas crudas para <i>V. parahaemolyticus</i> . entregadas por BGI.   | 22     |
| 5.1. Cantidad de lecturas crudas por muestra para <i>V. parahaemolyticus</i><br>entregadas por BGI. . . . .  | 34     |
| 5.2. Cantidad de lecturas eliminadas y obtenidas por muestra para <i>V.</i><br><i>parahaemolyticus</i> después de realizar la selección aleatoria. . . . . | 36     |
| 5.3. Cantidad de lecturas eliminadas por filtro realizado, y porcentaje de lecturas<br>eliminadas. . . . .   | 37     |
| 5.4. Coeficiente de determinación $R^2$ para los RPKM obtenidos con el alineador<br>BWA. . . . .   | 42     |
| 5.5. RPKM para cada uno de los sRNA por condición, y su expresión diferencial<br>utilizando los resultados obtenidos con el alineador BWA. . . . .         | 43     |
| 5.6. Resultados Expresión Diferencial utilizando los resultados del alineador BWA.   | 44     |
| 5.7. Cantidad de RNAs para <i>V. parahaemolyticus</i> . . . . .  | 46     |
| 5.8. Cantidad de pares de bases totales por cada tipo de RNA . . . . .   | 46     |
| 5.9. Expresión de sRNA en cada condición. . . . .  | 46     |
| 9.1. Cada línea de la sección de alineación del formato SAM posee las 11<br>columnas obligatorias descritas en la tabla. . . . .                           | 69     |
| 9.5. Valores de conteo con -a = 0 y -m = union para BWA . . . . .  | 76     |
| 9.6. Valores de conteo con -a = 3 y -m = union para BWA . . . . .  | 76     |
| 9.7. Valores de conteo con -a = 10 y -m = union para BWA . . . . .   | 76     |
| 9.8. Valores de conteo con -a = 0 y -m = intersection-nonempty para BWA . .  | 77     |
| 9.9. Valores de conteo con -a = 3 y -m = intersection-nonempty para BWA . .  | 77     |
| 9.10. Valores de conteo con -a = 10 y -m = intersection-nonempty para BWA . .  | 77     |
| 9.11. Valores de conteo con -a = 0 y -m = intersection-strict para BWA . . . . .   | 78     |
| 9.12. Valores de conteo con -a = 3 y -m = intersection-strict para BWA . . . . .   | 78     |
| 9.13. Valores de conteo con -a = 10 y -m = intersection-strict para BWA . . . . .  | 78     |
| 9.14. Valores de conteo con -a = 0 y -m = union para SMALT . . . . .   | 79     |
| 9.15. Valores de conteo con -a = 3 y -m = union para SMALT . . . . .   | 79     |
| 9.16. Valores de conteo con -a = 10 y -m = union para SMALT . . . . .  | 79     |

|   |    |
|---|----|
| 9.17. Valores de conteo con -a = 0 y -m = intersection-nonempty para SMALT .  | 80 |
| 9.18. Valores de conteo con -a = 3 y -m = intersection-nonempty para SMALT .  | 80 |
| 9.19. Valores de conteo con -a = 10 y -m = intersection-nonempty para SMALT   | 80 |
| 9.20. Valores de conteo con -a = 0 y -m = intersection-strict para SMALT . . .  | 81 |
| 9.21. Valores de conteo con -a = 3 y -m = intersection-strict para SMALT . . .  | 81 |
| 9.22. Valores de conteo con -a = 10 y -m = intersection-strict para SMALT . . .   | 81 |
| 9.23. RPKM para cada uno de los sRNA por condición, y su expresión diferencial<br>utilizando los resultados obtenidos con el alineador SMALT. . . . . | 83 |
| 9.24. Coeficiente de determinación $R^2$ para los RPKM obtenidos con el alineador<br>SMALT. . . . .   | 83 |
| 9.25. Resultados Expresión Diferencial utilizando los resultados del alineador<br>SMALT. . . . .  | 84 |
| 9.26. Listado de genes expresados diferencialmente para <i>V. parahaemolyticus</i> . .  | 94 |